

Weighing in on End Weight

Jason Grafmiller
Stanford University

Stephanie Shih
Stanford University
University of California, Berkeley

The Principle of End Weight maintains that constituents will occur in order of increasing weight (Behagel 1909; Quirk et al. 1985), but the precise definition of “weight” has been heavily debated. Previous proposals have defined weight as syntactic complexity (syntactic nodes or words), processing (dependencies), phonological complexity (lexical stresses), or phonological length (syllables). This paper presents a systematic investigation of these measures on constituent ordering in two constructions in spoken English. Our results show that the relative influences of phonological and syntactic weight measures vary across syntactic domains, and that weight effects cannot be reduced to a single dimension.

As approximations of syntactic complexity, counts of orthographic words or syntactic nodes have frequently been shown to be reliable predictors of constituent ordering (Wasow 2002; Szmrecsányi 2004; Bresnan and Ford 2010; a.o.). Phonological weight measures have also been suggested to affect constituent ordering (Benor and Levy 2006; McDonald et al. 1993; Anttila et al. 2010; Selkirk 1984; Zec and Inkelas 1990). However, these studies neither controlled for the influences of non-phonological predictors (see esp. Bresnan et al. 2007), nor did they account for the high degree of correlation between weight measures—e.g. the number of stresses increases as word count increases—which potentially masks the independent effects of individual variables.

In this study we examined the influences of various weight measures in genitive and dative construction choice (*the car’s wheel* ~ *the wheel of the car*; *give the dog the bone* ~ *give the bone to the dog*) in spoken American English using two relatively novel statistical techniques: information-theoretic model averaging (Burnham and Anderson 2004) and conditional random forests analysis (Strobl et al. 2009). Random forests analysis is robust to collinear data and is capable of detecting the independent importance of each variable. On the other hand, model averaging rejects the assumption of a single “best” model and instead allows us to make inferences about evidence from multiple candidate models with varying sets of parameters.

We studied five measures of weight, including the number of syntactic nodes, words, lexical stresses, syllables, and discourse-new referents in the constituents of genitive and dative constructions, while controlling for other known predictors of construction choice (Shih et al. 2009; a.o.). In both constructions, we found syntactic node count to be a highly predictive weight measure, but the reliability of other weight measures varied across the two constructions. Primary stresses were more reliable in predicting the genitive construction, while word and referent count were more reliable in the datives. We propose that this variation is evidence of differing constraints on organization within the NP and VP domains. More importantly, our study reveals that each of the measures investigated makes a distinct contribution to the choice of construction, thus challenging the common practice of reducing weight effects to a single syntactic or phonological variable. Finally, we note that while the two analytical methods used differ in important ways, there is a striking consistency between them, and we advocate this complementary approach to the study of linguistic variables.

Select references

- Anttila, Arto, Matthew Adams, and Michael Speriosu. 2010. "The role of prosody in the English dative alternation." *Language and Cognitive Processes*.
- Behagel, O. 1909. "Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern." *Indogermanische Forschungen*. 25: 110-42.
- Benor, Sarah Bunin and Roger Levy. 2006. "The Chicken or the Egg? A Probabilistic Analysis of English Binomials." *Language*. 82(2): 233-278.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. "Predicting the Dative Alternation." In Bouma, G., I. Kraemer, and J. Zwarts (ed). *Cognitive Foundations of Interpretation*. Royal Netherlands Academy of Science. 69-94.
- Bresnan, Joan and Marilyn Ford. 2010. "Predicting syntax: Processing dative constructions in American and Australian varieties of English." *Language*. 86(1): 168-213.
- Burnham, Kenneth P. and David R. Anderson. "Multimodel inference." *Sociological Methods and Research*. 33(2): 261-304.
- Comrie, Bernard. 2003. "On explaining language universals." In Tomasello, M. (ed). *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Erlbaum. 195-210.
- Gibson, Edward. 1998. "Linguistic Complexity: locality of syntactic dependencies." *Cognition*. 68: 1-76.
- Gibson, Edward. 2000. "The dependency locality theory: A distance-based theory of linguistic complexity." In Miyashita, Y., A. Marantz, and W. O'Neil (ed). *Image, Language, Brain*. Cambridge, MA: MIT Press. 95-126.
- Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- McDonald, Janet L., Kathryn Bock, and Michael H. Kelly. 1993. "Word and World Order: Semantic, Phonological, and Metrical Determinants of Serial Position." *Cognitive Psychology*. 25: 188-230.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.
- Selkirk, Elisabeth O. 1984. *Phonology and Syntax: the Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Shih, Stephanie, Jason Grafmiller, Richard Futrell, and Joan Bresnan. 2009. "Rhythm's role in genitive and dative construction choice in spoken English." Paper presented at the 31st annual meeting of the Linguistics Association of Germany (DGfS). University of Osnabrück, Germany. March 4, 2009.
- Strobl, Carolin, James Malley, and Gerard Tutz. 2009. "An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging and random forests." *Psychological Methods*. 14(4): 323-348.
- Szmrecsányi, Benedikt. 2004. "On Operationalizing Syntactic Complexity." *Journées internationales d'Analyse statistique des Données Textuelles*. 7: 1031-38.
- Wasow, Tom. 2002. *Postverbal Behavior*. Stanford, CA: CSLI Publications.
- Zec, Draga and Sharon Inkelas. 1990. "Prosodically Constrained Syntax." In Inkelas, S. and D. Zec (ed). *The Phonology-Syntax Connection*. Stanford, CA: Center for the Study of Language and Information.